中文题目：数据治理背景下文献计量与专利分析的比较性研究

薛皓白

深圳大学城图书馆　深圳　518055

摘要：[目的/意义] 为了利用论文和专利两种数据源进行联合分析，从多维度评价科研成果、追踪前沿主题或研究科学与技术的协同关系。 [方法/过程] 本文对文献计量和专利分析中的基本概念和常用方法进行了梳理和归类，对每类组合的异同之处进行了深入的比较和分析，给出了产生这些差异的根本原因，并总结出了进行论文-专利联合分析的基本方法和注意事项。[结果/结论] 研究表明，文献计量关注学术成果的影响力而专利分析关注技术领域的发展竞争态势及其商业价值，因此两者在分析方法和内容上各有侧重；文献计量学基本定律和常用的文献计量工具同样适用于专利分析，但需要针对专利文献的特点进行适当的调整；论文-专利联合分析可以通过研究论文和专利在研究领域、机构作者、引文和主题词四个方面的关联来实现。

关键词：文献计量；专利分析；比较研究；联合分析；论文-专利关联

一、　引言

期刊论文和专利同属科学知识的载体，两者包含了全球绝大多数的科技情报，是世界上最重要的科技信息源[1]。利用统计学的方法去定量分析这些知识载体，能够使之具有纵览全局或预测的功能，这种方法一般被称之为文献计量或专利分析[2-4]。但期刊论文和专利本质上属于不同的文献类型：前者是探讨学术问题、分享研究成果的工具和手段；而后者是记载发明内容、保护申请人权利的法律文件。因此两者无论在内容形式，还是目的用途上均存在显著的差异。事实上，针对期刊论文已经出现了如 InCites、CiteSpace 和 VOSviewer 等文献分析工具，并逐渐发展出文献计量学等相关研究领域[5, 6]。而针对专利文献，也已经出现了如 DII（Derwent Innovations Index）、Innography 和 Incopat 等专利分析工具，同时涌现出大量的有关专利分析的专著[2, 7]。但由于期刊论文和专利本质上的不同，两者所涉及的基本概念和常用方法存在一定的差异，而深入分析、比较这些概念和方法之间的差异，是进行论文-专利联合分析的基础。联合分析不仅可以对科技成果进行多维度的评价、追踪前沿研究主题，而且可以揭示科学与技术之间的协同关系，具有非常重要的意义。

*本文系深圳市图书情报科研项目“数据治理背景下文献计量与专利分析的比较性研究”（项目编号：深文图情 2022210）研究成果之一。

作者简介：　薛皓白(ORCID：0000-0002-8062-6155)，助理研究员，博士，E-mail：xuehb@utszlib.edu.cn

分析、比较文献计量和专利分析之间的异同，有助于促进两者之间的相互借鉴和协同发展。但目前鲜有相关的研究报道，绝大多数的研究仅限于文献计量和专利分析的简单组合。例如，不同学者已针对立体显示[8]、LED[9]、基因测序[10]、风力发电[11]、区块链[12]等研究领域展开了文献计量和专利分析。但上述研究均仅限于先利用文献分析工具统计论文，再利用专利分析工具统计专利，并没有将二者有机统一起来进行协同分析。高继平[13]、赵辉[14]和 Fred[15]等人分析了期刊论文与专利的关联，研究了科学与技术的互动并提出了基于论文-专利联合分析的研究前沿分析方法，但这些研究没有对文献计量和专利分析中涉及的基本概念和常用方法进行系统的梳理、分析和比较。

本文对文献计量和专利分析中的基本概念和常用方法进行了梳理和归类，对每类组合的异同之处进行了深入的比较和分析，给出了产生这些差异的根本原因，并总结出论文-专利联合分析的基本方法和注意事项。本文是进行文献-专利联合分析的基础，也是研究多维度科研评价、前沿主题追踪和研究科学技术协同关系的基础。

二、 比较与分析

论文和专利本质上属于不同的文献类型，因此文献计量和专利分析的目的和用途之间也存在一定的差异。其中，文献计量一般用于学术研究、科学评价和学术出版等领域，是学者、研究机构和学术出版机构制定研究计划和评定学术成果的重要工具；而专利分析则一般用于技术研究、市场研究和竞争情报等领域，是企业制定专利战略和技术布局的重要工具。因此，文献计量和专利分析的侧重点有所差异，前者更关注学术研究的成果和影响力，而后者则更关注技术领域的发展、竞争态势及其商业价值。但在具体的数据分析方法和报告内容格式上，两者之间又存在很大的相似性。一般来讲，文献计量和专利分析均包括趋势分析、构成/排序分析、关联分析、引文分析、聚类分析和主题演化分析等几种分析方法。

(一) 趋势分析

趋势分析研究某一机构、个人或研究领域的论文发表数量或专利申请/授权数量随时间的变化规律，揭示其发展阶段和内在原因，并对未来的发展趋势进行预测。其中，论文的趋势分析通常以出版年为依据，而专利的申请趋势和授权趋势分析则分别以申请日和授权日为依据。趋势分析有时会和构成分析或排序分析

一同使用。

文献计量学的相关研究表明，科学文献的数量按指数规律增长。各国专利机构每年公布出版的专利文献也呈指数上升趋势。但若分析对象仅限于某一研究主题或技术领域，则一般认为增长规律为逻辑型，分为诞生、发展、成熟、分化（或衰落）几个阶段[3]。在专利分析中，该方法又被称为技术生命周期分析[2]，以便决策者在不同的技术生命周期阶段制定与之相适应的技术发展策略，是专利分析的重点内容之一。常见的技术生命周期分析方法有专利指标法、相对增长率法、技术生命周期（TCT）图法、TCT 计算法和 S 曲线数学模型法等[2]。

(二) 构成和排序分析

构成和排序分析都是先从不同的角度将研究对象分为若干类，然后分析比较各个类的比例和数量。其中构成分析侧重于比较不同类的比例，一般以饼图、细胞图的形式呈现；而排序分析则侧重于比较不同类的相对数量，一般以柱状图、气泡图的形式呈现，两者可以相互结合使用。根据分析角度和分类依据的不同，一般可以将论文和专利按照学科与研究领域、机构与作者、地域与国家以及法律状态进行分类研究。

1. 学科与研究领域

以科睿唯安公司为例，其 Web of Science 和 JCR 数据库将期刊划分为约 250 个研究领域，而 ESI 数据库则将期刊划分为 22 个学科。利用 InCites 数据库则可以对不同学科或研究领域的论文数量和比例进行统计分析。由于每个学科或研究领域均包含若干种期刊，因此若要对研究对象进行更为微观的分析，则可以按照期刊名对论文进行分类。类似地，由于每种期刊均包含若干不同的研究主题，而关键词则可以很好地反映这些研究主题，因此可以利用关键词对论文做更加微观的分类和统计。

与此相对的是，专利没有期刊名、关键词等著录项，而是直接依据专利著录项中的分类信息（如国际专利分类 IPC、联合专利分类 CPC 等）对其技术领域进行分类和统计。以国际专利分类 IPC 为例，其从高到低包括部、大类、小类、大组和小组五个层级，其中大类和小类大致相当于论文中的学科或研究领域，小类和大组大致相当于论文中的来源出版物（期刊名），大组和小组则大致相当于论文中的关键词。近年来，德温特世界专利索引数据库（DWPI）的分类代码（DC）

和手工代码（MC）通过技术专家人工标引的方式建立起一个覆盖所有技术领域、分层级且基于字母和数字的标引系统，以帮助突出专利的创新点和应用特点，是与 IPC 完全不同的分类体系。

文献计量学中的布拉德福定律表明若将期刊按照其所刊登某学科或研究领域的论文数量三等分为核心区、相关区和外围区，则三个区的期刊数量满足$1:a:a^2$的关系，而这种分布关系同样存在于专利文献中[16,17]。将布拉德福定律应用于国际专利分类号（IPC）进行区域划分，能够有效地确定某技术领域的核心分类，从而为核心技术的寻找提供了理论依据[16,17]。但有证据表明，若以专利的引文数量为分区依据，则引文分布较布拉德福定律的预测结果更为集中、紧凑[18]。

将论文学科分类与专利技术分类进行映射连接的过程被称为类目映射关系分析法，其为论文-专利联合分析的方法之一。例如，英国工程技术学会（IET）的 Inspect 数据库能够直接根据输入的 IPC 专利分类号检索到相关的论文，从而节省了专利审查员和研究人员的时间[19,20]。Verbeek 等通过研究专利 IPC 分类与引文期刊学科分类之间的映射关系，建立了技术创新与科学研究之间的关联[21]。目前，类目映射法的理论方法研究较为欠缺、关联结果的准确性较差、相关研究较少[14]。

2. 机构与作者

大部分论文和专利数据库都能够直接对论文的所属机构（或专利的申请人）和作者（或发明人）进行排序、构成分析。文献计量学中的洛特卡定律揭示了作者比例$f(x)$与文献数量x之间的平方反比规律，即$f(x)=C/x^2$ [3]。研究表明专利文献中同样存在类似的分布规律，但其分布曲线较论文更为陡峭，即$f(x)=C/x^k$（$k>2$）[1]。因此科学生产率集中于少数研究人员的规律对论文和专利文献同样适用[1,22,23]。另外，专利分析中通常会对申请人的类型进行分析，如企业、科研机构或大学等，但论文分析中很少对作者机构的类型进行统计。

此外，由于科技类论文通常由不同机构和作者合作完成，因此 CiteSpace 等文献计量软件能够对作者、机构和国家\地区的合作网络进行可视化分析，图中用气泡的大小表示机构论文数量的多少，用气泡间联系的粗细表示机构间合作次数的多寡。大部分专利数据库虽然也能够对申请人或发明人之间的合作进行可视化分析，但一般是通过对技术功效矩阵（或其它二维构成矩阵）进行自定义设置

得到的，其可视化效果有限，使用频次也较低。例如，Innography 数据库提供了一种 Custom 4D 热力图的功能，若将横、纵坐标均设置为发明人，则可以看到发明人之间合作的情况。

对论文合作网络的研究表明，研究人员合作者数量分布呈现幂律分布的特征，即大部分科研人员仅与少数合作者共同研究，而极少数科学家则积累了大量的合作者[24, 25]。另一项研究表明，专利合作的广泛程度不如期刊论文，即专利合作主要以机构内的合作为主，而期刊论文则可能由多个机构的研究人员共同完成[14]。

作者-发明人关系分析法是进行论文-专利联合分析的方法之一，该方法通过对既发表论文又申请专利的学者（即学术型发明人）的论文发表和专利申请情况来研究科学和技术之间的关联性。学术型发明人通常处于其所在学科的核心位置，是研究合作的桥梁和媒介，推动了科技的进步和发展[26]。该方法的关键和难点在于解决论文作者和专利发明人的同一性识别问题，另外可用于统计分析的数据也较为缺乏[27]。

3. 地域与国家

InCites 数据库能够按照地域对论文进行排序、统计分析。不同地域的论文与专利的数量反映了其在科学与技术领域的生产能力，有研究表明上述指标与该地域的经济实力（如 GDP）成正比，而与该地域的面积、人口和其他指标无关[1]。CiteSpace 还能够对不同地域间的合作进行可视化，从而反映出国际间的合作关系。需要注意的是，论文分析中的地域是指作者所属机构的地理位置，而专利分析中的地域则较为复杂。这是因为专利的审查和授权是以国家为单位的，即某国授予的专利仅在该国有效。因此专利分析中需要明确专利的起源国和目标国，起源国可以根据申请人地址来判断，而目标国则可以根据专利文献的公开地域（公开号中的地域代码）来判断[7]。图表的表现形式既可以采用分别对来源国和目标国进行分析的形式，也可以采用技术输入输出矩阵的形式同时对来源国和输出国的构成和关联进行分析[2]。

4. 法律状态

分析专利的法律状态可以衡量竞争对手的技术研发实力和技术领域的专利活跃度，因此专利的法律状态也是分析的重点之一。我国的专利类型分为发明、实用新型和外观设计三种，其中发明专利的审查包括初步审查和实质审查两个阶

段[28]。初步审查通过后公开的专利文献号以 A 结尾，实质审查通过后公开的专利文献号以 B 结尾[7]。因此发明专利申请公开后未必授权，专利分析中一般用“专利授权率”来测算专利申请的技术质量。此外，发明、实用新型和外观设计分别有 20 年、10 年和 15 年的有效期限，因此专利分析中一般用“专利存活率”来测算专利的法律有效状态。专利类型、授权率和存活率都是专利分析中的重点内容。

由于专利具有严格的地域性，本国申请人的发明创造若要获得其他国家的保护，就必须分别向其他国家申请专利，由此产生的一组内容相同或基本相同的专利出版物被称之为专利族，其中每件专利互为同族专利，通过优先权这一媒介联系在一起[7]。专利族又可以分为简单专利族、复杂专利族、扩展专利族、人工专利族等多种形式[7]。同族专利的数量不仅反映了专利申请地域的广泛程度，也反映了该专利的重要性和潜在经济价值。

此外，专利的转让、权利人变更、诉讼、复审决定、质押、许可、标准必要专利等法律信息反映了专利的经济价值和重要性，因此也是专利分析的重点内容。而上述这些有关法律状态的分析与研究在论文分析中是不存在的。

(三) 关联分析

关联分析主要用来研究两个变量之间的相互关联，其可以视为上节构成分析或排序分析的二维延申，即二维构成分析和二维排序分析。例如，竞争情报中常用的 SWOT 分析法和波士顿矩阵就是一种二维构成分析，即将研究对象同时按照两个指标分为四个象限进行分类研究[29]。若二维图表的横、纵坐标为连续的统计指标而非离散的分类项，例如横轴是专利的数量而纵轴是专利的质量（用引用量表示），则该图表可以视为一种二维排序分析方法。关联分析在专利分析或竞争情报研究中十分常见，但在论文分析或文献计量学研究中则较为少见，这可能是和专利涉及技术、经济和法律等很多方面有关，即单一指标难以反映创新主体参与市场竞争的各个方面。

1. 二维构成分析

二维构成分析的横、纵坐标均为分类项。例如，若纵坐标表示专利的来源国，横坐标表示专利的目标国，圆圈的大小表示满足条件专利的数量，则该二维构成图就是上文“地区与国家”一节提及的技术输入输出矩阵。再如，若横、纵坐标

均为关键词，圆圈的大小表示同时出现横、纵坐标关键词论文的数量，则该二维构成图实质上反映了关键词之间的共现情况，文献计量学中又称之为共词分析。专利分析中也有类似的“共词分析”，但其纵坐标一般为技术词，其反映了专利所采用的技术手段；横坐标一般为功效词，其反映了专利所实现的技术效果。因此，该二维构成图表又被称为技术功效矩阵，突出展示了专利所采用的技术手段以及所实现的技术效果，是专利分析中的重要内容之一[2]。

我国专利法规定，专利的保护对象是技术方案，而技术方案的判断依据就是采用了技术手段，解决了技术问题，并实现了技术效果。由此可见，专利严格的撰写规则和语法结构为技术功效矩阵的实施提供了便利。利用技术功效矩阵，能够很清晰地看出某种技术手段所能够实现的技术效果，或者要实现某种技术效果所需要采用的技术手段，从而为技术空白点、技术研发热点和突破点的寻找提供了帮助。因此，技术功效矩阵与共词分析在功能作用上有所区别。除了技术功效矩阵以外，将技术领域作为纵坐标，同时将地域、申请人或发明人作为横坐标的二维构成分析也非常常见。这是因为上述数据均可通过分析软件直接获得，无需人工处理，因此制图较为简便。同时该二维构成分析能够清楚地看出不同地域、申请人和发明人在不同技术领域的积淀和布局。

2. 二维排序分析

二维排序分析的横、纵坐标均为连续的统计指标。例如，若同时从专利强度（反映现有专利的数量和质量）和专利申请增长率两个指标对不同的技术领域进行分析，则可以帮助企业发现那些同时具有较高技术份额和吸引力的领域，从而指导企业的研发投资行为。再如，若同时从专利活动度（专利数量）和专利质量两个指标对不同的企业进行分析，则可以将企业划分为技术领导者、技术竞争者、技术活跃者和技术落后者四种类型，从而确立技术创新的跟踪对象和优先选择的技术合作伙伴。此外，也可以将专利指标与非专利指标结合来对分析对象进行二维排序，例如可以同时从企业财务和专利实力两个角度对合作对象进行排序。

3. 与竞争情报的联系

二维构成和二维排序在竞争情报分析中也获得了广泛的应用[29]。例如，SWOT 分析法同时从企业的内部资源和外部环境两个角度将不同因素划分为优势、劣势、机遇和威胁四种情况，从而使企业战略计划制定更加科学全面。情景

分析法则同时从战略重要性和发生概率两个角度对不同的情境进行模拟演习，并据此建立预警系统或制定相关战略。波士顿矩阵根据市场增加率和相对市场占有率两个指标将产品划分为四种类型，从而帮助企业据此制定不同的策略[29]。由于竞争情报分析与专利分析之间存在着密切的联系，因此关联分析在专利分析中应用广泛。与此相对的是，关联分析在文献计量中应用较少，因此可以考虑将上述方法引入文献计量，例如可以同时从论文的数量和质量，或者引文的数量和增长率等不同角度对作者和机构进行分析。

(四) 引文分析

引文即论文或专利在撰写过程中所引用的其他有关文献。论文引用的对象被称为"参考文献"（cited reference），而引用该论文的对象则被称之为"引证文献"（citation）。与之相对的是，专利引证的对象被称为"后向引证"（backward citation）或"引用专利（cited patent）"，而引证该专利的对象则被称之为"前向引证"（forward citation）或"施引专利"（citing patent）。上述对应概念间除术语略有差异之外，含义基本相同。但需要注意的是，专利引文除了包括说明书中的引用参考文献之外，还包括审查员添加的审查对比文件，这些文件可以在专利扉页或检索报告中的审查对比文件目录中找到[2, 7]。因此，在 Derwent 专利数据库中，被发明人和审查员引用的专利和论文被细分为四类分别进行统计分析。考虑到发明人可能会故意隐藏最为密切相关的技术以换取专利的授权与保护，由审查员添加的后向引证通常被认为更具客观性。

论文的引文数量遵循二八定律，即大多数论文很少被引用，但少数论文的引用则要多很多。普赖斯认为引用分布近似于一个幂指数为 3 的幂律函数[30]，但进一步的研究表明论文分布更符合对数正态分布，且归一化后的分布曲线适用于所有的研究领域[31]。对专利文献的研究表明，美国专利同样符合类似的分布规律[1]。

专利引用不仅代表着发明的重要性，而且还反映了该专利的经济影响力[32]。例如，拥有高被引专利的产品会获得更多的需求而公司则会获得更高的市值[33]。更为重要的是，专利的商业价值与它的引用正相关，而且高度非线性。例如有研究表明，拥有 14 次引用的专利，其估价是获得 8 次引用专利的 100 倍[34]。

1. 分析方法

与其他文献一样，专利除了直接引用之外，也存在耦合和共被引的现象。其

中，耦合是指两篇文献同时引用了第三篇文献，而共被引是指两篇文献同时被第三篇文献引用。因此，耦合分析多用来探寻学科基础，而共被引分析则多用来追踪研究前沿。此外，耦合和共被引分析均可用于对文献进行聚类分析，但使用不如基于共词分析的聚类分析普遍。

通过计量专利和论文之间相互引用从而分析科学与技术之间相互作用的方法称为引文关联分析法[14]。由于引用是建立在对被引文献消化、吸收、再利用的基础之上的，因此引文关联分析能够反映出科学和技术之间的相互转换和推动作用，有助于识别出具有商业价值的研究前沿[14]。但该方法存在专利-论文互引数量稀少、获取难度大、引用动机不明确和时效性较差等缺点。

2. 引文指标

研究表明，论文和专利的被引用次数可以在一定程度上反映其重要程度，因此两者均发展出一系列基于引文数量的评价指标。例如，影响因子（IF）[35]和 H-指数[36]可以分别用来对期刊和学者的学术影响力进行定量评价，而 ESI 数据库则可以根据引用次数对国家、机构和学科进行排名。此外，ESI 还分别基于引文的数量和突现度提出了 ESI 高被引论文和热点论文等指标，用来对重点论文进行筛选。专利分析中同样具有基于引文数量的统计指标，常见的有当前影响指数（CII）和科学关联度（SL）两种，均由美国 CHI 公司提出[4]。其中，当前影响指数（CII）指的是分析对象每项专利平均被引用的次数与所在行业每项专利平均被引用次数的比值，比值大于一则说明分析对象的专利影响力高于行业平均水平。当前影响指数（CII）与 InCites 中的学科规范化引文影响力（CNCI）指标的含义和用法类似。科学关联度（SL）则是专利分析中特有的指标，定义为每项专利平均引用学术论文的数量，其反映了专利与科学技术关联的紧密程度[4]。该指标可用于分析科学与技术之间的互动关系。

值得注意的是，很多专利数据库并不只依赖引用指标来筛选重点专利，这是因为专利具有经济、技术和法律三重属性，因此重点专利的衡量指标通常会根据分析目的对多个指标进行加权平均。例如，Innography 数据库中的“专利强度”是专利价值判断的综合指标，其涵盖了权利要求数量、引用与被引用次数、是否涉案、专利时间跨度、同族专利数量等多个因素，可以用来对核心专利进行快速筛选[2]。

(五) 聚类分析

聚类分析是数据挖掘中的重要方法，其能够在无需人工标注和预先训练的条件下揭示出数据背后隐含的信息，从而突破了现有的知识框架，避免了仅凭经验和专业知识进行分类导致的误区。聚类分析一般包括文本聚类分析和引证聚类分析两种，其中前者是最主要、最常见的聚类分析手段。文本聚类对论文或专利的文本信息（标题、摘要、关键词、权利要求书等）进行聚类，从而揭示出某一学科领域的研究进展和研究前沿，或某一技术领域的专利布局、发展态势和空白点。因此主题词共现，也就是共词分析是进行文本聚类分析的前提，也是论文分析的核心内容，因为其反映了不同主题词之间的关联度，从而为研究领域的划分奠定了基础。主题词在论文中的分布规律符合齐普夫定律[3]，而直接验证专利文献中关键词分布规律的文献较少，现仅知专利中停用词分布大致符合齐普夫定律[37]。

由于聚类分析涉及大量复杂的计算，因此通常需要借助数据库或软件等专业工具实现。论文分析中，CiteSpace、HistCite、VOSviewer 都是常见的聚类分析工具。以 CiteSpace 为例，其以节点间连线的粗细表示关键词共现的数量。若某节点与其它节点间的连线越多、越粗，其也就越靠近图谱的核心位置[6]。CiteSpace 就是根据节点位置和连线情况来确定聚类中心和类别的，并用方框和标签加以区别。此外，CiteSpace 通常用不同颜色的圆环和连线来表示关键词出现的时间，时间越新则颜色越鲜艳，因此其生成的聚类地图能够展示出不同研究领域的新旧情况[6]。以上这些聚类分析工具同样可以用于专利的分析，但应当注意把专利的文本格式转换为软件能够识别的标准格式[7]。例如，专利文献没有关键词信息，因此首先需要通过语义分析工具（例如 Patentics[38]）对专利文本进行分词，再进行文本格式的转换。

除了上述通用的聚类分析工具之外，绝大多数的专利数据库都配置有自己的聚类分析功能，从而对用户选定的专利文件进行一键聚类，并通过多种形式的聚类地图进行展示。例如，Innography、incoPat 和智慧芽等专利数据库均提供了类似专利地图的表现形式，其中山峰的高度代表专利的密度，相隔的距离代表技术领域的差异，部分数据库还能以不同颜色的点表示不同属性的专利[2, 4]。但整体而言，专利地图的聚类展示效果与 CiteSpace 等通用工具实现的聚类效果类似。此外，Innography 和 incoPat 等专利数据库还能通过多环图（即双层饼图）或树

状图的形式实现类似层次聚类的效果，从而帮助分析人快速提炼技术点。

以主题词为连接点，分析其在论文和专利中的共现情况，从而判断论文和专利的主题关联性，进而分析科学与技术之间的关联关系的方法被称之为主题词关联分析法[14]。由于论文和专利用语的差异性较大，因此在实践中通常是先分别对论文和专利聚类之后，再进行跨数据源的主题比较[14]。比较方法可以采用相似性计算，也可以由人工对资料研读后进行判断。主题词关联分析法的适用性较好，能够对多种数据源进行联合分析，但其揭示的结果可能缺乏真实的互动关系。

(六)主题演化分析

主题演化分析在专利分析中通常被称为技术路线图，而在文献计量中则经常被称为时区图和时间线图，两者在本质上都是主题演化分析的具体方式。技术路线图是专利分析中的一项重要战略规划和决策工具，其能够基于专利文献信息描绘某技术领域的主要技术发展路径和关键技术节点，从而帮助企业厘清技术的发展主流，获取更多的竞争情报并把握技术未来发展方向[2, 7]。技术路线图通常以年代作为横坐标，以技术构成（技术要素）或者技术功效作为纵坐标，并在对应的位置标识出专利数量或专利号。为了表现出专利间的技术关联性，技术路线图有时会以连线的形式表示出专利间的引证关系或共现关系。目前的技术路线图通常以技术发展需求为主线，综合多种因素筛选出关键技术节点，从而使技术路线更接近实际。

主题演化分析可以采用不同的研究方式，例如既可以根据文献引证路径进行年代切分，以研究不同阶段的技术沿革情况；也可以利用文本聚类后的关键词分析研究主题的变化和关联情况。其中后者在论文分析中获得了广泛的应用，例如CiteSpace 提供的时间线图或时区图就可以视为技术路线图的一种[6, 39]。在时间线图中，相互平行的时间线代表不同的聚类类别，时间线上不同位置的圆圈代表不同的关键词，其中圆圈的位置表示该关键词首次出现的年份。因此，时间线图反映了各个研究领域内不同研究主题首次出现的时间，进而揭示了该领域的发展脉络、热点转变和研究前沿，具有重要的参考价值。此外，该图由文献计量软件自动生成，无需任何人工操作，其便捷性和客观性也是其获得广泛应用的重要原因之一。最后，论文分析中有时还会利用突现词分析的方法来追踪研究热点和前沿[6]，但该方法在专利分析中较为罕见。

三、 结论

本文对文献计量和专利分析中的基本概念和常用方法进行了梳理和归类，并对每类组合的异同之处进行了深入的比较和分析，给出了产生这些差异的原因，总结出了进行论文-专利联合分析的常用方法和注意事项。从本文的研究中可以得到以下基本结论：

1. 文献计量和专利分析的目的和侧重点有所差异。文献计量更关注学术研究的成果和影响力，因此构成/排序分析、共现分析和引文分析是其研究的重点；而专利分析则更注重技术领域的发展竞争态势及其商业价值，因此趋势分析、关联分析和主题演化分析是其研究的重点，另外专利的法律分析和价值评估也是其独有的重要组成部分。

2. 文献计量学的基本定律一般认为适用于专利分析，但不能完全将论文的拟合参数用于专利文献，而是需要针对专利文献的特点进行适当的调整和解释。通用的文献计量工具一般也可以用于专利分析，但需要对专利文献的格式进行标准化或进行适当的预处理。

3. 文献计量和专利分析所使用的方法、工具和表现形式大体相似，仅在部分术语、统计指标和图表类型上有所差异，专利分析的图表和指标种类略多于文献计量。论文-专利联合分析可以通过类目映射关联、作者-发明人关联、引文关联和主题词关联四种方式实现，以研究科学与技术的互动作用。

四、 参考文献

[1] NARIN F. Patent bibliometrics [J]. Scientometrics, 1994, 30(1): 147-155.

[2] 马天旗. 专利分析——方法，图表解读与情报挖掘 [M]. 北京: 知识产权出版社, 2015.

[3] 邱均平. 文献计量学 [M]. 北京: 科学出版社, 2019.

[4] 文庭孝. 专利信息计量学 [M]. 北京: 科学出版社, 2017.

[5] 侯剑华. 知识计量与可视化——原理、方法及应用 [M]. 北京: 科学出版社, 2016.

[6] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for information Science and Technology, 2006, 57(3): 359-377.

[7] 马天旗. 专利分析——检索、可视化与报告撰写 [M]. 北京: 知识产权出版社, 2021.

[8] 殷媛媛，肖沪卫. 基于论文专利的科学技术互动发展趋势研究——以立体显示产业为例 [J]. 情报杂志, 2011, 30(06): 25-29+81.

[9] 许晓阳. 基于论文和专利的研究前沿识别方法研究——以 LED 产业为例 [D]; 中国科学技术信息研究所, 2014.

[10] 孙轶楠，杜建，唐小利. 基于科学论文和专利信息的基因测序技术领域创新态势分析 [J]. 中华医学图书情报杂志, 2017, 26(05): 31-37.

[11] 侯剑华，都佳妮. 科学—技术互动视角下战略性新兴技术演进路径研究——以风力发电机技术为例 [J]. 科技与经济, 2015, 28(01): 12-16.
[12] 张维冲，王芳，赵洪. 多源信息融合用于新兴技术发展趋势识别——以区块链为例 [J]. 情报学报, 2019, 38(11): 1166-1176.
[13] 高继平，丁堃，滕立, et al. 专利—论文混合共被引网络下的知识流动探析 [J]. 科学学研究, 2011, 29(08): 1184-1189+1146.
[14] 赵辉. 基于论文-专利关联的前沿主题分析方法研究 [D]. 北京; 中国科学院大学(中国科学院文献情报中心), 2020.
[15] FRED Y Y，HUANG M-H，CHEN D-Z. Comparative Study of Trace Metrics between Bibliometrics and Patentometrics [J]. Journal of Data and Information Science, 2017, 1(2): 13-31.
[16] 吕义超，刘红光，王君. 布拉德福定律在专利文献中应用的可行性研究 [J]. 图书情报研究, 2011, 4(02): 49-52.
[17] 张鹏，刘平，唐田田, et al. 布拉德福定律在专利分析系统中的应用 [J]. 现代图书情报技术, 2010, (Z1): 84-87.
[18] HUANG M H，HUANG W T，CHANG C C, et al. The greater scattering phenomenon beyond B radford's law in patent citation [J]. Journal of the Association for Information Science and Technology, 2014, 65(9): 1917-1928.
[19] VERBANDT Y，VADOT E. Non-patent literature search at the European Patent Office [J]. World Patent Information, 2018, 54: S72-S77.
[20] 刘闯. 英国工程技术学会(IET)用户培训 [EB/OL]. http://calis20.calis.edu.cn/Home/Menu/4b3e3b02-67f5-4760-8f86-ee157e39d5e7, 2023-05-05.
[21] VERBEEK A，DEBACKERE K，LUWEL M, et al. Linking science to technology: Using bibliographic references in patents to build linkage schemes [J]. Scientometrics, 2002, 54: 399-420.
[22] 冯瑶，解烈军，董其军. 宁波市发明专利生产率的洛特卡分布研究 [J]. 科技情报开发与经济, 2013, 23(01): 125-127.
[23] 杨中楷，林德明，刘佳. 洛特卡定律适用于专利文献的再验证 [J]. 图书情报工作, 2013, 57(09): 92-95.
[24] NEWMAN M E. Coauthorship networks and patterns of scientific collaboration [J]. Proceedings of the national academy of sciences, 2004, 101(suppl_1): 5200-5205.
[25] WANG D，BARABáSI A-L. The science of science [M]. Cambridge University Press, 2021.
[26] BALCONI M，BRESCHI S，LISSONI F. Networks of inventors and the role of academia: an exploration of Italian patent data [J]. Research policy, 2004, 33(1): 127-145.
[27] 张俊. 科学与技术的关联性研究 [D]; 南京农业大学, 2014.
[28] 知识产权出版社有限责任公司. 专利代理师资格考试指南 [M]. 北京: 知识产权出版社, 2021.
[29] 王知津. 竞争情报 [M]. 北京: 科学技术文献出版社, 2005.
[30] PRICE-DE SOLLA D. Networks of Science Papers [J]. Science, 1965, 149.
[31] RADICCHI F，FORTUNATO S，CASTELLANO C. Universality of citation distributions: Toward an objective measure of scientific impact [J]. Proceedings of the National Academy of Sciences, 2008, 105(45): 17268-17272.
[32] JAFFE A B，FOGARTY M S，BANKS B A. Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation [J]. The Journal of Industrial

Economics, 1998, 46(2): 183-205.
[33] TRAJTENBERG M. A penny for your quotes: patent citations and the value of innovations [J]. The Rand journal of economics, 1990: 172-187.
[34] HARHOFF D，NARIN F，SCHERER F M, et al. Citation frequency and the value of patented inventions [J]. Review of Economics and statistics, 1999, 81(3): 511-515.
[35] GARFIELD E. Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies [J]. Science, 1972, 178(4060): 471-479.
[36] HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National academy of Sciences, 2005, 102(46): 16569-16572.
[37] BLANCHARD A. Understanding and customizing stopword lists for enhanced patent mapping [J]. World Patent Information, 2007, 29(4): 308-316.
[38] Patentics [EB/OL]. https://www.patentics.com/, 2023-05-05.
[39] MORRIS S A，YEN G，WU Z, et al. Time line visualization of research fronts [J]. Journal of the American society for information science and technology, 2003, 54(5): 413-422.

英文题目： A comparative study of the bibliometrics and patentometrics in the context of data governance

Haobai Xue

University Town Library of Shenzhen　Shenzhen　518055

Abstract: [Purpose/significance] In order to utilize the data source of papers and patents for joint analysis, to evaluate the scientific achievements from multiple perspectives, to track the research frontiers and to study the mutual relationship between science and technology. [Method/process] This paper sorted out the concepts and methods commonly used in the bibliometrics and patentometrics, thoroughly compared and analyzed their similarities and differences, provided a reasonable explanation for these differences and summarized the basic methods and key points for paper-patent joint analysis. [Result/Conclusion] The research shows that the bibliometrics emphasize the influence of academic achievements whereas the patentometrics emphasize the evolvement and competition of technical fields as well as their commercial value, therefore the focus of analysis contents of bibliometrics and patentometrics are different. Basic laws and software in bibliometrics can also be applied to patentometrics but should be adjusted to account for the characteristics of patents. The paper-patent joint analysis can be achieved by studying the connections of papers and patents in research area, institutes or authors, citations and key words.

Key words: bibliometrics; patent analysis; patentometrics; comparative study; joint analysis; paper-patent connection